Title
Mesolimbic dopamine release conveys causal associations

Journal
Science, 378(6626)

ISSN
0036-8075

Authors
Jeong, Huijeong
Taylor, Annie
Floeder, Joseph R
et al.

Publication Date
2022-12-23

DOI
10.1126/science.abq6740

Peer reviewed

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2023 June 23.

Published in final edited form as:

Science. 2022 December 23; 378(6626): eabq6740. doi:10.1126/science.abq6740.

Mesolimbic dopamine release conveys causal associations

Huijeong Jeong[1], **Annie Taylor**[2,†], **Joseph R Floeder**[2,†], **Martin Lohmann**[4], **Stefan Mihalas**[4,5], **Brenda Wu**[1], **Mingkang Zhou**[1,2], **Dennis A Burke**[1], **Vijay Mohan K Namboodiri**[1,2,3,*]

[1]Department of Neurology, University of California, San Francisco, CA, USA

[2]Neuroscience Graduate Program, University of California, San Francisco, CA, USA

[3]Weill Institute for Neuroscience, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA, USA

[4]Allen Institute for Brain Science, Seattle, WA, USA

[5]Department of Applied Mathematics, University of Washington, Seattle, WA, USA

Abstract

Learning to predict rewards based on environmental cues is essential for survival. It is believed that animals learn to predict rewards by updating predictions whenever the outcome deviates from expectations, and that such reward prediction errors (RPEs) are signaled by the mesolimbic dopamine system—a key controller of learning. However, instead of learning prospective predictions from RPEs, animals can infer predictions by learning the retrospective cause of rewards. Hence, whether mesolimbic dopamine instead conveys a causal associative signal that sometimes resembles RPE remains unknown. We developed an algorithm for retrospective causal learning and found that mesolimbic dopamine release conveys causal associations but not RPE, thereby challenging the dominant theory of reward learning. Our results reshape the conceptual and biological framework for associative learning.

How do animals learn to associate environmental cues with delayed outcomes such as rewards? It is widely believed that they learn a prospective prediction of how often reward follows a given cue. A simple way to learn such prospective predictions is to update one's prediction every time the outcome following a cue deviates from the prediction (Fig 1A, B). Such violations of reward predictions are commonly called reward prediction errors (RPEs). The simplest model in this family is the Rescorla-Wagner model (1). Temporal

*Corresponding author: VijayMohan.KNamboodiri@ucsf.edu.

Author contributions:

H.J. and V.M.K.N. designed the study. H.J., V.M.K.N., A.T., and M.L. performed simulations with input from S.M. H.J., M.Z., and V.M.K.N. set up instrumentation for behavior control and photometry. H.J., J.R.F., B.W., and D.A.B. performed experiments. H.J. performed analysis. H.J., A.T., and V.M.K.N. wrote the paper with help from all authors. V.M.K.N. supervised all aspects of the study.

†These authors contributed equally to this work (co-second author)

Competing interests: Authors declare that they have no competing interests.

Data and materials availability:

All data from this study are publicly available on NIH DANDI at https://dandiarchive.org/dandiset/000351

The codes for analysis and simulation are available publicly on Github (https://github.com/namboodirilab/ANCCR) and on Zenodo (77).

difference reinforcement learning (TDRL) models extend the Rescorla-Wagner model to account for cue-outcome delays and is the most widely accepted model of reward learning (2, 3). To account for delays, these models typically propose that a sequential pattern of neural activities ("states") tiles temporal delays and propagates predictions from the cue to the reward (Fig 1B). TDRL RPE has been successful at explaining the activity dynamics of dopaminergic cell bodies and release in the nucleus accumbens (4-13). Hence, TDRL RPE has become the dominant theory of dopamine's role as the critical regulator of behavioral learning (14-17).

An alternative approach to learn cue-reward associations is to infer the cause of meaningful outcomes such as rewards (18-20) (Fig 1A, C). Because causes must precede outcomes, a viable approach to infer whether a cue causes reward is to learn whether the cue consistently *precedes* reward. Predicting the future is highly demanding in a cue-rich environment but inferring the cause of a rarer meaningful outcome simply requires a memory of previous experience. If an animal knows that a stimulus it just received is meaningful (e.g., a reward), it can look back in memory to infer its cause. Given the central role of dopamine in learning, we hypothesized that dopamine may guide retrospective causal learning instead of conveying RPEs. Though the differences between prospective and retrospective learning may not be apparent at first glance, we show that these models make highly divergent predictions about mesolimbic dopamine dynamics. Here, we directly test between these models of the role of dopamine in associative learning.

A retrospective causal learning algorithm:

While some stimuli are innately meaningful, others acquire meaning after learning that they cause other meaningful stimuli (e.g., a cue that predicts reward becomes meaningful). We denote stimuli whose cause should be learned by the animal as "meaningful causal targets" and propose that mesolimbic dopamine signals whether a current event is a meaningful causal target (Figs 1C, S1, S2). We propose a causal inference algorithm that infers whether a cue is a cause of reward by measuring whether it precedes the reward more than that expected by chance (Figs 1C, S2), then converting this to a prospective prediction signal using Bayes' rule (Fig S3) (Supplementary Note 1), and finally using the net contingency between a cue and reward to build a cognitive map of causal associations (20) (Figs 1C, S4).

We developed this algorithm to address problematic temporal assumptions that are foundational to common conceptions of TDRL, which result in a non-scalable representation of time (21). We tested whether this new algorithm learns causal relationships without loss of generality across timescales. Consistent with this and unlike TDRL, our algorithm learns the underlying causal structure of a variety of complex environments across two orders of magnitude of timescales and explains well-established behavioral observations of the timescale invariance of learning (Figs S5, S6). The algorithm proposes that meaningful causal targets are signaled by an adjusted net contingency for causal relations (ANCCR, read "anchor") (Fig S4). The ANCCR-based causal learning model is consistent with simulations of classical results supporting the RPE coding hypothesis including dopaminergic responses to reward magnitude and probability, blocking, unblocking, overexpectation, conditioned inhibition, and trial-by-trial update of action probabilities (Fig 2). It is also consistent with

the observation that apparent negative RPEs in dopamine response are not as strong as positive RPEs of the same magnitude, even without assuming a floor effect in dopamine responses. Therefore, we reasoned that mesolimbic dopamine release has been tested only under conditions in which the ANCCR and RPE hypotheses make similar predictions, and that dopamine release may convey ANCCR instead of RPE.

In most behavioral tasks, prospective and retrospective associations are highly correlated and difficult to separate. To distinguish between the two hypotheses (RPE or ANCCR signaling by mesolimbic dopamine release), we performed eleven experimental tests. To maximize our ability to distinguish the models for strong inference (22), we designed the experiments such that the predictions of the two hypotheses are qualitatively different and often opposing. Because it has been proposed that distinct dopaminergic systems exist in the midbrain and that only some faithfully signal RPE (23-30), we tested these predictions by optically measuring sub-second mesolimbic dopamine release in the nucleus accumbens core (NAcc), a projection widely believed to encode RPE and shown to mediate Pavlovian learning (8-12, 31-33) (though see (34)) (Figs 3A, S7). We did so in mice using fiber photometry of the dopamine sensor dLight 1.3b expressed in NAcc (7, 35).

Tests 1 and 2 (unpredicted rewards):

We first tested between the two hypotheses in a simple experiment with divergent predictions. We presented naïve head-fixed mice with no experience in any laboratory behavior task with random unpredicted drops of a 15% sucrose solution delivered with an exponential inter-reward interval (IRI) distribution (mean = 12 s), while recording mesolimbic dopamine release in NAcc. In this task, the timing of individual sucrose deliveries cannot be anticipated based on the previous delivery, but the average rate of sucrose delivery is fixed (once every 12 s on average). Because the animal is experimentally naïve with no history of receiving sucrose prior to the onset of the experiment, the RPE hypothesis predicts high dopamine response to sucrose during the early exposures. This is because the sucrose is highly unpredicted initially. With repeated exposure to the context, the RPE is predicted to decrease slightly as the context becomes a predictor of the rewards. More formally, the internal IRI “states” in TDRL acquire positive value with experience (see Supplementary Note 2 for a consideration of a semi-Markov state space in TDRL (36)). Since RPE is the difference between the value of sucrose and the value of the IRI state that preceded sucrose delivery, RPE will reduce at sucrose delivery with repeated experience (Fig 3B, C).

On the other hand, the ANCCR hypothesis predicts that the response to sucrose will increase with repeated experience. This is because the predicted sucrose response is proportional to the difference between the average rate of previous sucrose deliveries calculated at sucrose delivery (including the current sucrose delivery) and the baseline average rate of previous sucrose deliveries (Fig 3B). Because both of these quantities are initially low in naïve animals that have no experience with sucrose, ANCCR of sucrose is low early in this task. ANCCR eventually reaches an asymptote of ~1 times the incentive value of sucrose (Methods) because the rate of sucrose calculated just prior to a sucrose delivery (i.e., excluding the current sucrose) is equal to the baseline average rate of sucrose. Thus, the

RPE hypothesis predicts that the dopamine response to sucrose will decrease over repeated experiences, while the ANCCR hypothesis predicts that the response will increase. Testing these differential predictions formed Test 1 (Fig 3B, C).

Observed mesolimbic dopamine release was consistent with ANCCR but not RPE (Fig 3D, E). Every animal showed an increasing sucrose response that reached a high positive asymptote. This is entirely inconsistent with RPE: because RPE is the difference between received and predicted reward, it cannot be higher than that for an unpredicted reward. These results also cannot be explained by RPE based on a slower learning of the incentive value of sucrose; animals actively licked to consume sucrose at high rates starting from the first delivery, demonstrating that sucrose had high value (Fig 3D, Fig S8). Such high motivation for sucrose from the onset of the experiment is consistent with well-known results that sugar is innately rewarding to mice (37). We also ruled out alternative hypotheses such as stress (Supplementary Note 3, Fig S8) or a non-specific increase in responses to the consummatory action (lick bout onset) (Fig S8).

We next tested a “trial-by-trial” prediction in this experiment by measuring the correlation between the dopamine response to a sucrose delivery and the previous IRI. Getting the next reward sooner than predicted would produce a larger RPE than getting the next reward later. Hence, the RPE hypothesis predicts a negative correlation between the dopamine response to a sucrose delivery and the previous IRI (36) (Fig 3B, F) (Supplementary Note 4). However, ANCCR predicts a positive correlation because the ANCCR of reward involves the subtraction of the baseline reward rate. Because the baseline reward rate declines with longer IRI, ANCCR should increase with longer IRI (Fig 3B, F). This was Test 2.

The experimentally observed correlation between dopamine response to sucrose and the previous IRI was positive, thereby being consistent with ANCCR but not RPE. We also ruled out the hypothesis that this positive correlation is simply due to an inability of animals to learn the mean IRI. This is because 1) the correlation was consistently positive for more than 800 experiences of sucrose (8 sessions) (Fig S8), 2) mice learn the average IRI within at most two sessions (Fig S8), 3) rodents can be as fast as Bayesian ideal observers in detecting changes in the rate of exponentially scheduled rewards (38), and 4) even the original experiments that inspired the Rescorla-Wagner model showed that animals learn the mean inter-reinforcer interval despite unpredictable timing (39, 40) (see (41) for a detailed discussion).

Tests 3-7 (Cue-reward learning):

Next, we studied dopamine response dynamics during cue-reward learning. We measured behavioral learning using anticipatory licking prior to the delivery of sucrose 3 s following onset of an auditory cue. Anticipatory licking reflects the prediction of upcoming reward across species, and this paradigm has provided some of the strongest support for TDRL RPE coding (4, 5, 42-45). During cue-reward learning, both RPE and ANCCR predict that dopamine responses to the cue will be low early in learning and high late in learning. Thus, the increase in dopamine response to cue can be used as a measure of dopaminergic learning (defined as dopaminergic signaling related to the external cue-reward association).

The RPE hypothesis predicts a tight relationship between the dynamics of behavioral and dopaminergic learning (Fig 4A). This is because TDRL RPE updates the value signal used for behavioral learning, and dopaminergic signaling in NAcc is necessary for the learning of anticipatory licking in head-fixed mice (32). On the other hand, the ANCCR of the cue is a continuously evolving estimate of whether the cue is itself a meaningful causal target due to its association with reward, and hence, is not predicted to evolve in lockstep with the behavior. Indeed, in the ANCCR hypothesis, associations are learned first, and then timing is learned: behavioral learning requires the threshold crossing of ANCCR to learn a causal model of the world ("cue causes reward"), followed by the separate learning of the temporal delay between cue and reward ("cue causes reward at a 3 s delay"). Only then does a timed decision signal for behavior become available (Fig 4B, S2). Thus, the ANCCR hypothesis predicts that the gradual dopaminergic learning of the cue response will significantly precede behavioral learning, and that behavioral learning will be much more abrupt than dopaminergic learning since it requires an internal threshold crossing of the net contingency between cue and reward (Test 3) (Supplementary Note 5). The observed dopaminergic dynamics during learning were consistent with ANCCR but not RPE: dopamine response to CS+ was evident long before animals showed anticipatory licking (Figs 4B-F, S9). In fact, dopamine cue responses were at their peak by the time of behavioral acquisition (Fig S10).

Further, when a learned delay between cue onset and reward (3 s) is extended permanently to a new, longer delay (9 s), RPE predicts that as animals learn the longer delay and suppress anticipatory licking at the previous short delay, there will be a concomitant reduction in the dopamine cue response due to temporal discounting (46). On the other hand, ANCCR predicts little to no change in the dopamine cue response as the structure of the task is largely unchanged (**Test 4**, Figs 4 G, S9, S10; intuitively, relative to the long intertrial interval, the cue-reward delay is still short). Experimentally, we observed that while the animals learned the new delay rapidly, dopaminergic cue response showed no significant change (Fig 4 G-I). After the extension of the cue-reward delay, RPE predicts a suppression of dopamine after the old delay expires without reward. Because the increase in cue-reward delay is permanent (unlike in prior experiments (45)), ANCCR predicts that the delay representation in the internal causal model of the animal would be updated to reflect the new delay. This predicts no reward omission response at the old delay (3 s) after the increase in the delay to 9 s. Thus, ANCCR predicts no negative omission response after the old delay expires without reward. (Test 5). Experimentally, we observed no suppression of dopamine response at 3 seconds in this experiment but did observe suppression in a separate experiment when the reward was indeed omitted (Figs 4J, S10).

Next, we tested extinction of a learned cue-reward association. Extinction of a learned association does not cause unlearning of the original association (47). Yet, TDRL learns a zero cue value following extinction, thereby predicting that the dopaminergic cue response will reduce to zero concomitant with behavioral learning. However, ANCCR includes the measurement of a retrospective association between the cue and reward. This association does not update without rewards and hence, does not degrade due to extinction. This "long-term memory" was observed previously in orbitofrontal neurons projecting to the

ventral tegmental area, the region where the somata of the mesolimbic dopamine neurons reside (19). Hence, the ANCCR hypothesis predicts that dopamine response will remain significantly positive long after animals learn to suppress anticipatory licking. This is because the cue remains a meaningful causal target despite extinction, even though the animals can learn extinction by noting that the base rate of rewards in the context becomes zero. Thus, Test 6 was whether dopamine cue response remained positive long after extinction was behaviorally learned (Fig 4J-L). As predicted by ANCCR but not RPE, dopamine cue response remained significantly positive well after animals cease to behaviorally respond to the cues (Fig 4J-L), consistent with prior studies (48, 49).

To test whether the significant positive dopamine responses following extinction reflect a retrospective association between the cue and reward, we selectively reduced the retrospective association without reducing the prospective association. We maintained the fixed reward following the cue but added unpredictable rewards during the intertrial interval. In this experiment, not all rewards are preceded by the cue (i.e., retrospective association is weak), but all cues are followed by reward (i.e., prospective association is high). ANCCR predicts a rapid drop in dopamine cue response, but RPE predicts no change in cue response if TDRL only considers the cue-reward "trial period" (**Test 7**, Fig S10). The dopamine cue response remained significantly positive but decayed across trials faster than during extinction (Fig 4M-P).

Test 8 ("trial-less" cue-reward learning):

We performed another test related to the temporal scalability of TDRL versus retrospective causal inference (**Test 8**, Fig 5). A key motivation for developing our model was that current TDRL models do not have a scalable representation of time, and hence fail to learn the correct structure of even simple environments in which a cue predicts a reward at a fixed delay with 100% probability (Fig S6). We devised an experiment in which a single cue predicted the reward at a fixed delay with 100% probability, but the cue occurred unpredictably with an exponentially distributed inter-cue interval between 0-99 s. We reduced the cue duration to 250 ms to allow nearby occurrences of the cue to be separated in time and had a long trace interval (3 s) following cue offset until reward delivery. Animals learned the cue-reward association quickly in this modified "trial-less" task (Fig S11).

In this task, a cue will occasionally be presented during the wait from the previous cue to its associated reward (Fig 5A). If the "trial period" for cue-reward tasks is considered to be the interval between the cue and reward, the next "trial" can occasionally start *before* the previous trial is completed. During these "intermediate" cues, TDRL resets its prediction because it assumes a new trial has started without reward in the previous trial, thereby resulting in a negative RPE (i.e., the intermediate cue signals that the reward will now be further delayed; intuitively, the intermediate cue implies omission of reward after the previous cue). This results from the inability of TDRL to learn the correct structure of the task, which is that every cue occurrence causes a reward at a fixed delay (Supplementary Note 6).

On the other hand, ANCCR will learn that the intermediate cue is qualitatively similar to the previous cue because both predict reward, but due to a local increase in cue rate, ANCCR predicts a lower but positive response to the intermediate cue (Fig 5A, B). We did not observe any negative dopamine response to the intermediate cue regardless of how baseline was measured, and instead observed a positive but weaker response, consistent with ANCCR but not RPE (Figs 5C, D, S11).

Tests 9-11 (backpropagation within a trial):

A critical postulate of the TDRL RPE account is that dopamine responses drive value learning of the immediately preceding state. We tested three predictions of this central postulate that are each inconsistent with ANCCR. The first is that during the acquisition of trace conditioning, dopamine response systematically backpropagates from the moment immediately prior to reward to the cue onset (50) (**Test 9**, Fig 6A). Unlike TDRL RPE, ANCCR does not make such a prediction since delay periods are not broken into states in ANCCR. The second is that during sequential conditioning (cue1 predicts cue2 predicts reward), dopamine response first increases to cue2 and then increases to cue1 (**Test 10**, Fig 6C). ANCCR instead predicts that dopamine responses to both cues will increase together and later diverge when cue2 is learned to be caused by cue1. The third is that artificially suppressing dopamine release from cue2 to reward during sequential conditioning will prevent learning of cue1 responses (**Test 11**, Fig 6E-H). In contrast, suppressing cue2 response in ANCCR only prevents the learning of the cue1→cue2 association and does not prevent the learning of cue1 response.

We tested the first prediction using the animals that underwent the previous cue-reward learning. Our observations were not consistent with a backpropagating bump of activity and were instead consistent with an increase in cue response over trials of learning (Fig 6B) (see Supplementary Note 9 for potential reasons for discrepancy with a recent study). To test the second and third predictions, we performed sequential conditioning with an experimental group receiving inhibition of dopaminergic cell bodies from cue2 to reward, and a no-opsin control group that received the same laser but no inhibition of dopamine neurons. We measured NAcc dopamine release in both groups. The control group allowed us to test the dynamics of dopamine responses during sequential conditioning in the absence of dopamine neuron inhibition (i.e., the second prediction). Consistent with ANCCR, we experimentally found that cue2 and cue1 responses increased together early in learning prior to separating later in learning (Fig 6D). To test the third prediction, we first verified robust inhibition of mesolimbic dopamine release during the cue2→reward delay in the experimental group (~0.6 times the reward response on day 1 of conditioning) (Supplementary Note 10). With such strong inhibition, TDRL RPE predicted no behavioral learning in this experiment, and a strong negative cue1 dopamine response (Figs 6H, S12). In contrast, ANCCR predicted largely intact learning of cue1, but with slower behavioral learning and reduced cue1 response (see Supplementary Note 10 for explanation). Consistent with ANCCR, we observed that every experimental animal learned the task and that mesolimbic dopamine acquired positive responses to cue1 in all experimental animals (Fig 6I).

Discussion:

The dynamics of mesolimbic dopamine release in NAcc were inconsistent with TDRL RPE across a multitude of experiments but remain consistent with a causal learning algorithm. The algorithm proposed here operates by testing whether a stimulus precedes reward beyond that expected by chance and by converting this association to a prospective prediction (Supplementary Note 7). Using this prediction, the algorithm learns a causal map of associations, and signals whether a stimulus has become a meaningful causal target following such learning. Though our data are inconsistent with encoding of TDRL RPE by mesolimbic dopamine release, our framework is not inconsistent with prediction errors in general. Indeed, "prediction errors" related to event rates are a part of our framework (Supplementary Note 4).

The algorithm and results presented here provide a unified account of numerous published observations. Evidence across multiple species and brain regions shows that in addition to prospective associations, the brain stores memories of retrospective associations (19, 51, 52). Behavioral learning is also guided by retrospective associations (18, 53). Dopamine responses remain significantly positive even to fully predicted, delayed rewards (4, 46, 54-56). This is usually explained by appealing to an internal uncertainty about the delay (46) but occurs without any accounting of temporal uncertainty in our theory (Fig 2A). Consistent with our theory, a previous study observed no correlation between temporal uncertainty of an animal and the dopaminergic response to a fully predicted, delayed reward (57). Under some settings, dopamine reward responses during cue-reward conditioning have been observed to increase during initial learning, before decreasing back (54). While this observation is not consistent with RPE, it naturally results from our algorithm if the animal had no exposure to the reward in the experimental context prior to conditioning, as was the case (Fig S13). This might also explain why NAcc dopamine response to a predicted punishment might increase in some scenarios, while the responses to repeated punishments at fixed intervals decrease (34) (punishments are also meaningful causal targets; see Supplementary Note 8). ANCCR also explains recent observations of dopamine ramps used in favor of the RPE hypothesis (58) (Fig S13). Our explanation is also consistent with dopamine ramps in the striatum reflecting a causal association between an action and reward (59). Finally, dopamine responses guide learning in a way that sometimes violates the predictions of model-free TDRL (17, 60-63). Our proposal that the dopaminergic system conveys whether cues are meaningful causal targets, thereby promoting the learning of their causes, explains these results (Fig S13).

Our work raises several questions for which reports in the literature suggests answers. First, how is retrospective cue-reward information conveyed to the dopaminergic system? Prior work suggests that the orbitofrontal cortex is a source of this information (19) (Fig S14). Second, how do animals infer the appropriate timescales in the world? Currently, we simply assume that animals set the appropriate timescale of an environment based on knowledge of the inter-reward interval. As a more principled solution, recent work has suggested that multiple parallel systems with different time constants exist in the brain and can learn a timescale invariant representation of past time (64-67). Third, are there as-yet unknown state space assumptions that make TDRL RPE fit our data? We cannot rule

out all possible assumptions of TDRL state spaces because there is unlimited flexibility in assuming the state space used by animals, thereby making them currently unfalsifiable (though see Fig S15). In the absence of such falsifiable assumptions, our work demonstrates that the TDRL algorithm with conventional state space assumptions does not explain the dynamics of dopamine release in NAcc. Fourth, does dopamine release in regions other than NAcc signal RPE? As mentioned in the introduction, we studied dopamine release in NAcc precisely because it is the region with the strongest support for the RPE hypothesis. Considering the theoretical advantages of ANCCR compared to TDRL RPE in learning associations between rates of events (Fig S6, S15B), we believe that dopamine release in other regions might also be inconsistent with TDRL RPE; though, this remains to be tested. Finally, since it has been demonstrated that animal behavior and neural activity for even simple Pavlovian associations may be explained by the learning of causal cognitive maps (68-71), is all associative learning, including for action-conditional cognitive maps (56, 59, 72-76), the product of causal inference? This remains to be addressed. Collectively, our data demonstrate that mesolimbic dopaminergic signaling in NAcc is inconsistent with the dominant theory of TDRL RPE signaling and instead guides a causal learning algorithm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank S. Gu for suggesting that successor representation may relate to predecessor representation by Bayes' rule, an insight critical to this formulation. We thank A. Mohebi for advice on the initial photometry setup. We thank J. Berke, M. Frank, L. Frank, M. Andermann, K. Wassum, M. Brainard, M. Stryker, A. Nelson, V. Sohaal, M. Kheirbek, H. Fields, Z. Knight, H. Shouval, J. Johansen, A. Kepecs, A. Lutas, M. Hussain Shuler, G. Stuber, M. Howard, A. Mohebi, T. Krausz, R. Gowrishankar, I.Trujillo-Pisanty, J. Levy, J. Rodriguez-Romaguera, R. Simon, M. Hjort, Z. C. Zhou, V. Collins, T. Faust, M. Duhne, and other Nam lab members for comments on the general conceptual framework, experiments and/or the manuscript/figures.

Funding:

National Institute of Mental Health grant R00MH118422 (V.M.K.N.)

National Institute of Mental Health grant R01MH129582 (V.M.K.N.)

Scott Alan Myers Endowed Professorship (V.M.K.N.)

References and Notes

1. Rescorla RA, Wagner AR, A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory. 2, 64–99 (1972).
2. Niv Y, Schoenbaum G, Dialogues on prediction errors. Trends Cogn Sci. 12, 265–272 (2008). [PubMed: 18567531]
3. Niv Y, Reinforcement learning in the brain. Journal of Mathematical Psychology. 53, 139–154 (2009).
4. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N, Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature. 482, 85–88 (2012). [PubMed: 22258508]
5. Schultz W, Dayan P, Montague PR, A Neural Substrate of Prediction and Reward. Science. 275, 1593–1599 (1997). [PubMed: 9054347]

6. Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G, Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nat. Neurosci 14, 1590–1597 (2011). [PubMed: 22037501]
7. Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD, Dissociable dopamine dynamics for learning and motivation. Nature. 570, 65–70 (2019). [PubMed: 31118513]
8. Hart AS, Rutledge RB, Glimcher PW, Phillips PEM, Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. J. Neurosci 34, 698–704 (2014). [PubMed: 24431428]
9. Day JJ, Roitman MF, Wightman RM, Carelli RM, Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat Neurosci. 10, 1020–1028 (2007). [PubMed: 17603481]
10. Phillips PEM, Stuber GD, Heien MLAV, Wightman RM, Carelli RM, Subsecond dopamine release promotes cocaine seeking. Nature. 422, 614–618 (2003). [PubMed: 12687000]
11. Saddoris MP, Cacciapaglia F, Wightman RM, Carelli RM, Differential Dopamine Release Dynamics in the Nucleus Accumbens Core and Shell Reveal Complementary Signals for Error Prediction and Incentive Motivation. J. Neurosci 35, 11572–11582 (2015). [PubMed: 26290234]
12. Saddoris MP, Sugam JA, Stuber GD, Witten IB, Deisseroth K, Carelli RM, Mesolimbic Dopamine Dynamically Tracks, and Is Causally Linked to, Discrete Aspects of Value-Based Decision Making. Biological Psychiatry. 77, 903–911 (2015). [PubMed: 25541492]
13. Bayer HM, Glimcher PW, Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. Neuron. 47, 129–141 (2005). [PubMed: 15996553]
14. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH, A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci. 16, 966–973 (2013). [PubMed: 23708143]
15. Chang CY, Esber GR, Marrero-Garcia Y, Yau H-J, Bonci A, Schoenbaum G, Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat Neurosci. 19, 111–116 (2016). [PubMed: 26642092]
16. Tsai H-C, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K, Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science. 324, 1080–1084 (2009). [PubMed: 19389999]
17. Maes EJP, Sharpe MJ, Usypchuk AA, Lozzi M, Chang CY, Gardner MPH, Schoenbaum G, Iordanova MD, Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. Nat Neurosci. 23, 176–178 (2020). [PubMed: 31959935]
18. Gallistel CR, Craig AR, Shahan TA, Contingency, contiguity, and causality in conditioning: Applying information theory and Weber's Law to the assignment of credit problem. Psychol Rev. 126, 761–773 (2019). [PubMed: 31464474]
19. Namboodiri VMK, Otis JM, van Heeswijk K, Voets ES, Alghorazi RA, Rodriguez-Romaguera J, Mihalas S, Stuber GD, Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat. Neurosci 22, 1110 (2019). [PubMed: 31160741]
20. K Namboodiri VM, Stuber GD, The learning of prospective and retrospective cognitive maps within neural circuits. Neuron. 109, 3552–3575 (2021). [PubMed: 34678148]
21. Namboodiri VMK, How do real animals account for the passage of time during associative learning? Behav Neurosci. 136, 383–391 (2022). [PubMed: 35482634]
22. Platt JR, Strong Inference. Science. 146, 347–353 (1964). [PubMed: 17739513]
23. Heymann G, Jo YS, Reichard KL, McFarland N, Chavkin C, Palmiter RD, Soden ME, Zweifel LS, Synergy of Distinct Dopamine Projection Populations in Behavioral Reinforcement. Neuron. 105, 909–920.e5 (2020). [PubMed: 31879163]
24. Menegas W, Bergan JF, Ogawa SK, Isogai Y, Umadevi Venkataraju K, Osten P, Uchida N, Watabe-Uchida M, Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. eLife. 4, e10032 (2015). [PubMed: 26322384]

25. Menegas W, Akiti K, Amo R, Uchida N, Watabe-Uchida M, Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat Neurosci. 21, 1421–1430 (2018). [PubMed: 30177795]
26. Lammel S, Lim BK, Malenka RC, Reward and aversion in a heterogeneous midbrain dopamine system. Neuropharmacology. 76, 351–359 (2014). [PubMed: 23578393]
27. Lutas A, Kucukdereli H, Alturkistani O, Carty C, Sugden AU, Fernando K, Diaz V, Flores-Maldonado V, Andermann ML, State-specific gating of salient cues by midbrain dopaminergic input to basal amygdala. Nat Neurosci. 22, 1820–1833 (2019). [PubMed: 31611706]
28. Saunders BT, Richard JM, Margolis EB, Janak PH, Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. Nat Neurosci. 21, 1072–1083 (2018). [PubMed: 30038277]
29. Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD, Mesolimbic dopamine signals the value of work. Nat. Neurosci (2015), doi:10.1038/nn.4173.
30. Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A, Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision. Current Biology. 27, 821–832 (2017). [PubMed: 28285994]
31. Darvas M, Wunsch AM, Gibbs JT, Palmiter RD, Dopamine dependency for acquisition and performance of Pavlovian conditioned response. Proceedings of the National Academy of Sciences. 111, 2764–2769 (2014).
32. Yamaguchi K, Maeda Y, Sawada T, Iino Y, Tajiri M, Nakazato R, Ishii S, Kasai H, Yagishita S, A behavioural correlate of the synaptic eligibility trace in the nucleus accumbens. Sci Rep. 12, 1921 (2022). [PubMed: 35121769]
33. Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, Rudarakanchana N, Halkerston KM, Robbins TW, Everitt BJ, Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. Behavioural Brain Research. 137, 149–163 (2002). [PubMed: 12445721]
34. Kutlu MG, Zachry JE, Melugin PR, Cajigas SA, Chevee MF, Kelly SJ, Kutlu B, Tian L, Siciliano CA, Calipari ES, Dopamine release in the nucleus accumbens core signals perceived saliency. Current Biology. 31, 4748–4761.e8 (2021). [PubMed: 34529938]
35. Patriarchi T, Cho JR, Merten K, Howe MW, Marley A, Xiong W-H, Folk RW, Broussard GJ, Liang R, Jang MJ, Zhong H, Dombeck D, von Zastrow M, Nimmerjahn A, Gradinaru V, Williams JT, Tian L, Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science. 360, eaat4422 (2018). [PubMed: 29853555]
36. Daw ND, Courville AC, Tourtezky DS, Touretzky DS, Representation and timing in theories of the dopamine system. Neural Comput. 18, 1637–1677 (2006). [PubMed: 16764517]
37. Tan H-E, Sisti AC, Jin H, Vignovich M, Villavicencio M, Tsang KS, Goffer Y, Zuker CS, The gut-brain axis mediates sugar preference. Nature. 580, 511–516 (2020). [PubMed: 32322067]
38. Gallistel CR, Mark TA, King AP, Latham PE, The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. J Exp Psychol Anim Behav Process. 27, 354–372 (2001). [PubMed: 11676086]
39. Rescorla RA, Pavlovian conditioning and its proper control procedures. Psychol Rev. 74, 71–80 (1967). [PubMed: 5341445]
40. Rescorla RA, Probability of shock in the presence and absence of cs in fear conditioning. Journal of Comparative and Physiological Psychology. 66, 1–5 (1968). [PubMed: 5672628]
41. Gallistel CR, Robert Rescorla: Time, Information and Contingency. Revista de Historia de la Psicología. 42, 7–21 (2021).
42. K Namboodiri VM, Hobbs T, Trujillo-Pisanty I, Simon RC, Gray MM, Stuber GD, Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate. Current Biology. 31, 5176–5191.e5 (2021). [PubMed: 34637750]
43. Fiorillo CD, Newsome WT, Schultz W, The temporal precision of reward prediction in dopamine neurons. Nat. Neurosci 11, 966–973 (2008). [PubMed: 18660807]

44. Pastor-Bernier A, Stasiak A, Schultz W, Reward-specific satiety affects subjective value signals in orbitofrontal cortex during multicomponent economic choice. Proceedings of the National Academy of Sciences. 118, e2022650118 (2021).
45. Hollerman JR, Schultz W, Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci 1, 304–309 (1998). [PubMed: 10195164]
46. Kobayashi S, Schultz W, Influence of reward delays on responses of dopamine neurons. J. Neurosci 28, 7837–7846 (2008). [PubMed: 18667616]
47. Bouton ME, Maren S, McNally GP, Behavioral and neurobiological mechanisms of pavlovian and instrumental extinction learning. Physiological Reviews. 101, 611–681 (2021). [PubMed: 32970967]
48. Pan W-X, Schmidt R, Wickens JR, Hyland BI, Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. J. Neurosci 28, 9619–9631 (2008). [PubMed: 18815248]
49. Zhong W, Li Y, Feng Q, Luo M, Learning and Stress Shape the Reward Response Patterns of Serotonin Neurons. J. Neurosci 37, 8863–8875 (2017). [PubMed: 28821671]
50. Amo R, Matias S, Yamanaka A, Tanaka KF, Uchida N, Watabe-Uchida M, A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. Nat Neurosci 25, 1082–1092 (2022). [PubMed: 35798979]
51. Bouchard KE, Brainard MS, Neural Encoding and Integration of Learned Probabilistic Sequences in Avian Sensory-Motor Circuitry. J. Neurosci 33, 17710–17723 (2013). [PubMed: 24198363]
52. Komura Y, Tamura R, Uwano T, Nishijo H, Kaga K, Ono T, Retrospective and prospective coding for predicted reward in the sensory thalamus. Nature. 412, 546–549 (2001). [PubMed: 11484055]
53. Manzur HE, Vlasov K, Lin S-C, A retrospective and stepwise learning strategy revealed by neuronal activity in the basal forebrain (2022), p. 2022.04.01.486795, (available at https://www.biorxiv.org/content/10.1101/2022.04.01.486795v1).
54. Coddington LT, Dudman JT, The timing of action determines reward prediction signals in identified midbrain dopamine neurons. Nat. Neurosci 21, 1563–1573 (2018). [PubMed: 30323275]
55. Lee K, Claar LD, Hachisuka A, Bakhurin KI, Nguyen J, Trott JM, Gill JL, Masmanidis SC, Temporally restricted dopaminergic control of reward-conditioned movements. Nat. Neurosci 23, 209–216 (2020). [PubMed: 31932769]
56. Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, Witten IB, Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature. 570, 509–513 (2019). [PubMed: 31142844]
57. Hughes RN, Bakhurin KI, Petter EA, Watson GDR, Kim N, Friedman AD, Yin HH, Ventral Tegmental Dopamine Neurons Control the Impulse Vector during Motivated Behavior. Current Biology. 30, 2681–2694.e5 (2020). [PubMed: 32470362]
58. Kim HR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, Zhang Y, Li Y, Watabe-Uchida M, Gershman SJ, Uchida N, A Unified Framework for Dopamine Signals across Timescales. Cell. 183, 1600–1616.e25 (2020). [PubMed: 33248024]
59. Hamid AA, Frank MJ, Moore CI, Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. Cell. 184, 2733–2749.e16 (2021). [PubMed: 33861952]
60. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G, Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci. 20, 735–742 (2017). [PubMed: 28368385]
61. Sharpe MJ, Batchelor HM, Mueller LE, Yun Chang C, Maes EJP, Niv Y, Schoenbaum G, Dopamine transients do not act as model-free prediction errors during associative learning. Nature Communications. 11, 106 (2020).
62. Seitz BM, Hoang IB, DiFazio LE, Blaisdell AP, Sharpe MJ, Dopamine errors drive excitatory and inhibitory components of backward conditioning in an outcome-specific manner. Curr Biol. 32, 3210–3218.e3 (2022). [PubMed: 35752165]
63. Trujillo-Pisanty I, Conover K, Solis P, Palacios D, Shizgal P, Dopamine neurons do not constitute an obligatory stage in the final common path for the evaluation and pursuit of brain stimulation reward. PLoS One. 15, e0226722 (2020). [PubMed: 32502210]

64. Goh WZ, Ursekar V, Howard MW, Predicting the Future With a Scale-Invariant Temporal Memory for the Past. Neural Comput. 34, 642–685. [PubMed: 35026027]
65. Shankar KH, Howard MW, A scale-invariant internal representation of time. Neural Comput. 24, 134–193 (2012). [PubMed: 21919782]
66. Tsao A, Sugar J, Lu L, Wang C, Knierim JJ, Moser M-B, Moser EI, Integrating time from experience in the lateral entorhinal cortex. Nature. 561, 57–62 (2018). [PubMed: 30158699]
67. Wei W, Mohebi A, Berke JD, Striatal dopamine pulses follow a temporal discounting spectrum (2021), p. 2021.10.31.466705, (available at https://www.biorxiv.org/content/10.1101/2021.10.31.466705v2).
68. Madarasz TJ, Diaz-Mataix L, Akhand O, Ycu EA, LeDoux JE, Johansen JP, Evaluation of ambiguous associations in the amygdala by learning the structure of the environment. Nature Neuroscience. 19, 965–972 (2016). [PubMed: 27214568]
69. Gershman SJ, Niv Y, Exploring a latent cause theory of classical conditioning. Learn Behav. 40, 255–268 (2012). [PubMed: 22927000]
70. Balsam PD, Gallistel CR, Temporal maps and informativeness in associative learning. Trends Neurosci. 32, 73–78 (2009). [PubMed: 19136158]
71. Gershman SJ, Blei DM, Niv Y, Context, learning, and extinction. Psychol Rev. 117, 197–209 (2010). [PubMed: 20063968]
72. Syed ECJ, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME, Action initiation shapes mesolimbic dopamine encoding of future rewards. Nat Neurosci. 19, 34–36 (2016). [PubMed: 26642087]
73. Collins AL, Greenfield VY, Bye JK, Linker KE, Wang AS, Wassum KM, Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. Sci Rep. 6, 20231 (2016). [PubMed: 26869075]
74. Guru A, Seo C, Post RJ, Kullakanda DS, Schaffer JA, Warden MR, Ramping activity in midbrain dopamine neurons signifies the use of a cognitive map (2020), p. 2020.05.21.108886, (available at https://www.biorxiv.org/content/10.1101/2020.05.21.108886v1).
75. Hollon NG, Williams EW, Howard CD, Li H, Traut TI, Jin X, Nigrostriatal dopamine signals sequence-specific action-outcome prediction errors. Current Biology. 31, 5350–5363.e5 (2021). [PubMed: 34637751]
76. van Elzelingen W, Warnaar P, Matos J, Bastet W, Jonkman R, Smulders D, Goedhoop J, Denys D, Arbab T, Willuhn I, Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. Current Biology. 32, 1163–1174.e6 (2022). [PubMed: 35134325]
77. Jeong H, Taylor A, Floeder JR, Lohmann M, Mihalas S, Wu B, Zhou M, Burke DA, K Namboodiri VM, Mesolimbic dopamine release conveys causal associations, Version 1, Zenodo (2022); https://zenodo.org/record/7302777#.Y4j503bMI2w.
78. Gavornik JP, Shuler MGH, Loewenstein Y, Bear MF, Shouval HZ, Learning reward timing in cortex through reward dependent expression of synaptic plasticity. PNAS. 106, 6826–6831 (2009). [PubMed: 19346478]
79. Namboodiri VMK, Mihalas S, Marton TM, Hussain Shuler MG, A general theory of intertemporal decision-making and the perception of time. Front Behav Neurosci. 8, 61 (2014). [PubMed: 24616677]
80. Tobler PN, Fiorillo CD, Schultz W, Adaptive Coding of Reward Value by Dopamine Neurons. Science. 307, 1642–1645 (2005). [PubMed: 15761155]
81. Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB, Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci 19, 845–854 (2016). [PubMed: 27110917]
82. Akam T, Walton ME, pyPhotometry: Open source Python based hardware and software for fiber photometry data acquisition. Sci Rep. 9, 3521 (2019). [PubMed: 30837543]
83. Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, Deisseroth K, Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell. 162, 635–647 (2015). [PubMed: 26232229]
84. Martianova E, Aronson S, Proulx CD, Multi-Fiber Photometry to Record Neural Activity in Freely-Moving Animals. JoVE (Journal of Visualized Experiments), e60278 (2019).

85. Ludvig EA, Sutton RS, Kehoe EJ, Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. Neural computation. 20, 3034–3054 (2008). [PubMed: 18624657]

86. Pearl J, Causal inference in statistics: An overview. Statistics surveys. 3, 96–146 (2009).

87. Gallistel CR, Gibbon J, Time, rate, and conditioning. Psychol Rev. 107, 289–344 (2000). [PubMed: 10789198]

88. Ulrich-Lai YM, Christiansen AM, Ostrander MM, Jones AA, Jones KR, Choi DC, Krause EG, Evanson NK, Furay AR, Davis JF, Solomon MB, de Kloet AD, Tamashiro KL, Sakai RR, Seeley RJ, Woods SC, Herman JP, Pleasurable behaviors reduce stress via brain reward pathways. Proc. Natl. Acad. Sci. U.S.A 107, 20529–20534 (2010). [PubMed: 21059919]

89. Holly EN, Miczek KA, Ventral tegmental area dopamine revisited: effects of acute and repeated stress. Psychopharmacology (Berl). 233, 163–186 (2016). [PubMed: 26676983]

90. Stelly CE, Tritley SC, Rafati Y, Wanat MJ, Acute Stress Enhances Associative Learning via Dopamine Signaling in the Ventral Lateral Striatum. J Neurosci. 40, 4391–4400 (2020). [PubMed: 32321745]

91. George D, Rikhye RV, Gothoskar N, Guntupalli JS, Dedieu A, Lázaro-Gredilla M, Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. Nat Commun. 12, 2392 (2021). [PubMed: 33888694]

92. Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TEJ, The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell. 183, 1249–1263.e23 (2020). [PubMed: 33181068]

93. Hayden B, Niv Y, The case against economic values in the brain (2020), , doi:10.31234/osf.io/7hgup.

94. Etscorn F, Stephens R, Establishment of conditioned taste aversions with a 24-hour CS-US interval. Physiological Psychology. 1, 251–259 (1973).

95. Pearl J, Causal Diagrams for Empirical Research. Biometrika. 82, 669–688 (1995).

96. Bright IM, Meister MLR, Cruzado NA, Tiganj Z, Buffalo EA, Howard MW, A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. PNAS. 117, 20274–20283 (2020). [PubMed: 32747574]

97. 10.5281/zenodo.7302776

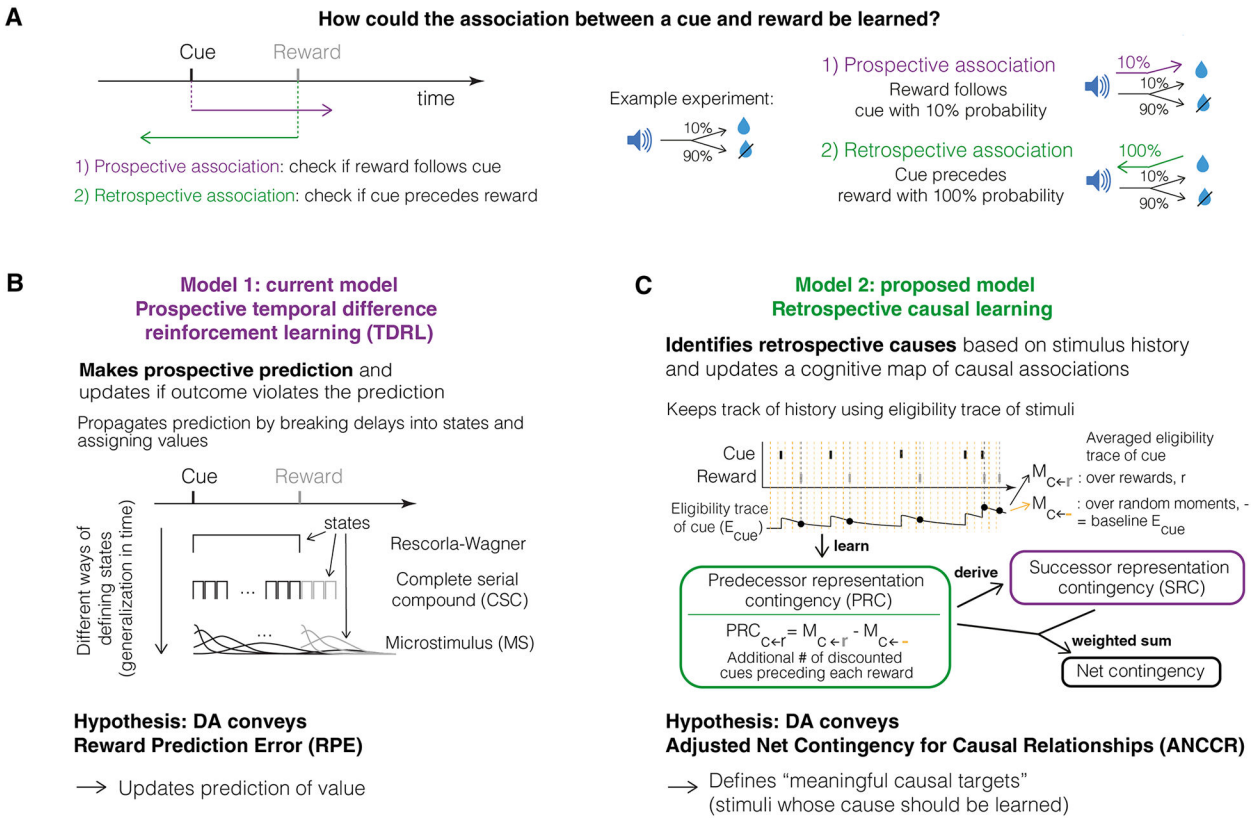

Fig. 1. An algorithm for uncovering causal associations in an environment.

A. Animals can learn cue-reward associations either prospectively (“does reward follow cue?”) or retrospectively (“does cue precede reward?”). **B.** The dominant model for cue-reward learning is temporal difference reinforcement learning, which learns the prospective association between a cue and reward, i.e., a measure of how often the reward follows the cue (cue value). To this end, the algorithm looks forward from a cue to predict upcoming rewards. When this prediction is incorrect, the original prediction is updated using a reward prediction error (RPE). The simplest of this family of models is the Rescorla-Wagner model which does not consider the delay between cue and reward. Temporal difference reinforcement learning (TDRL) algorithms extend this simple model to account for the cue-reward delay by modeling it as a series of states that measure time elapsed since stimulus onset. Two such examples are shown. **C.** Here, we propose an algorithm which retrospectively learns the causes of meaningful stimuli such as rewards (Fig S1-4). Because causes precede outcomes, causal learning only requires a memory trace of the past. In our mechanistic model, a memory trace of prior stimuli is maintained using an exponentially-decaying eligibility trace for a stimulus (78), which allows the online calculation of the experienced rate of this stimulus (79). We hypothesized that mesolimbic dopamine activity signals ANCCR, a quantity that allows measuring whether an experienced stimulus is a meaningful causal target.

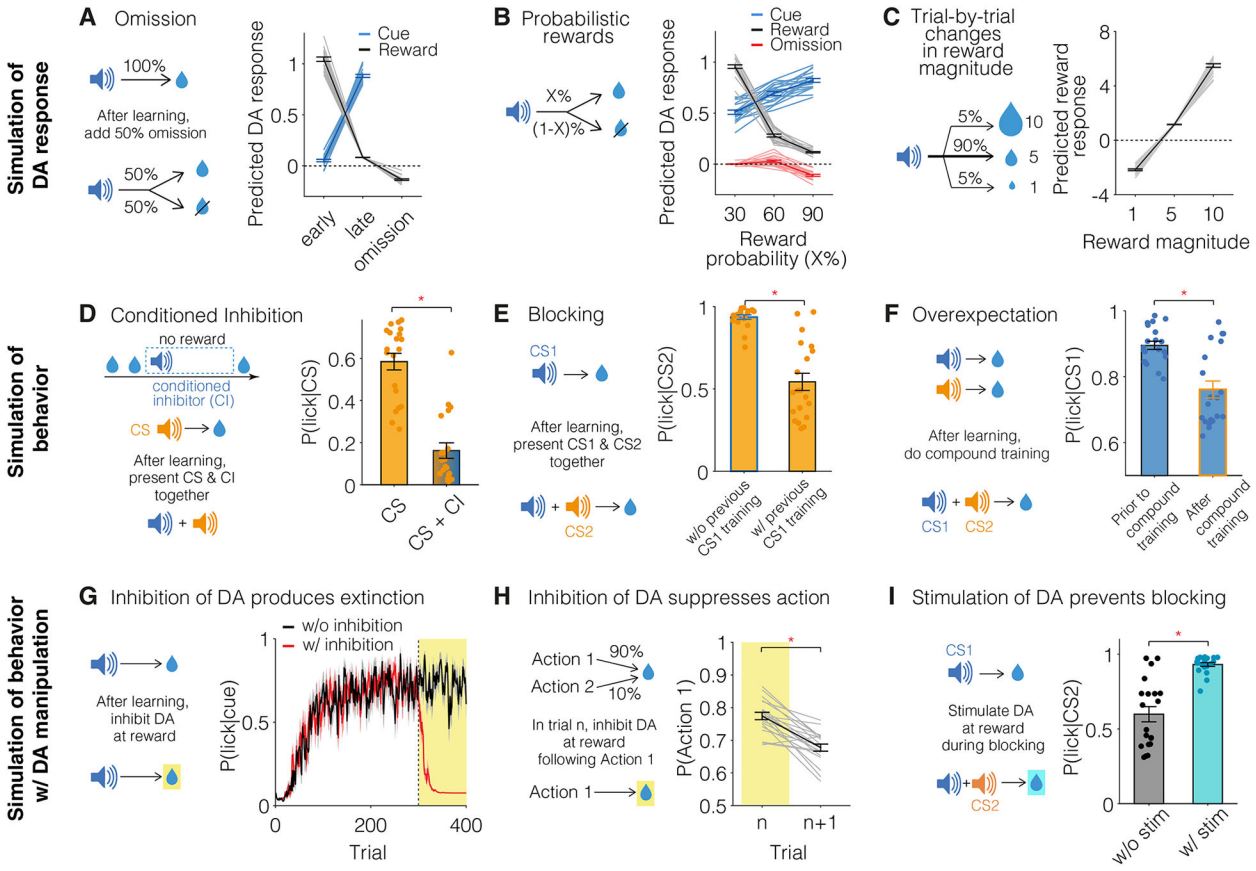

Fig. 2. The retrospective causal algorithm produces a signal similar to temporal difference reward prediction error (RPE) in simulations of previous experiments.

A. During simple conditioning of a cue-reward association, ANCCR appears qualitatively similar to an RPE signal, being low before and high after learning for the cue, whereas being high before and low after learning for the reward, and negative after omission of an expected reward. All error bars are standard error of the mean throughout the manuscript. **B.** For probabilistic rewards, ANCCR produces qualitatively similar responses as RPE for cue, reward, and omission. Note that in **B**, animals were never trained on a fully predicted reward. Slight differences in omission responses from **A** to **B** result from this difference. **C.** For trial-by-trial changes in reward magnitude, ANCCR produces reward responses similar to positive and negative RPEs (similar to (80)). **D-F.** Simulations of ANCCR learning produces behavior consistent with conditioned inhibition (**D**), blocking (**E**), and overexpectation **(F). G.** Simulated inhibition of dopamine at reward time in cue-reward conditioning produces extinction of learned behavior (similar to (55)). **H.** Simulation of dopamine inhibition at reward time produces trial-by-trial changes in behavior (similar to (81)). **I.** Simulation of unblocking due to dopamine activation at reward during blocking (similar to (14)).

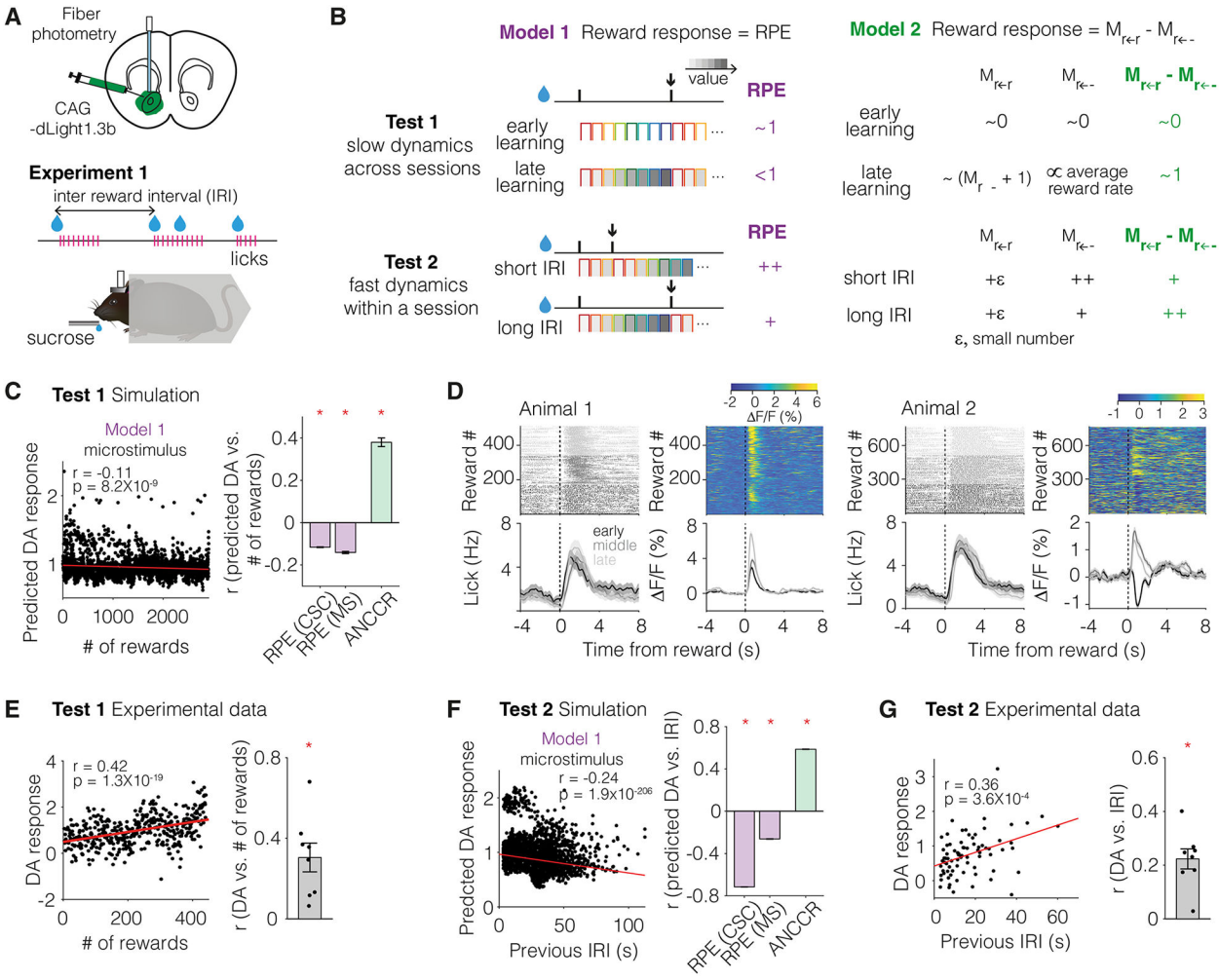

Fig. 3. The dynamics of dopamine responses to unpredicted rewards are consistent with ANCCR, but not TDRL RPE.

A. For the first two tests, we gave experimentally naïve mice random unpredictable sucrose rewards immediately following head-fixation while recording sub-second dopamine release in NAcc using the optical dopamine sensor, dLight 1.3b (Methods). Animals underwent multiple sessions with 100 rewards each (n=8 mice). **B.** Theoretical predictions for both models. **Test 1:** As a naïve animal receives unpredicted rewards, the RPE model predicts high responses since the rewards are unpredicted. Nevertheless, since the inter-reward interval (IRI) states acquire value over repeated experience, the RPE at reward will reduce with repeated experience. On the other hand, ANCCR predicts low reward responses early since an experimentally naïve animal will have no prior expectation/eligibility trace of sucrose early in the task but will subsequently approach a signal that is ~1 times the incentive value of sucrose. **Test 2:** The reward response following a short IRI will be larger in the RPE model because the reward was received earlier than expected, thereby resulting in a negative correlation between dopamine reward response and the previous IRI. However, since ANCCR has a subtractive term proportional to the baseline reward rate ($M_{r\leftarrow-}$ in the figure), and baseline reward rate reduces with longer IRI, ANCCR predicts a positive correlation between dopamine reward response and the previous IRI. **C.** Simulations

confirming the intuitive reasoning from **B** for Test 1. CSC and MS stand for complete serial compound and microstimulus, respectively. (one sample t-test against a null of zero; t(99) = RPE (CSC), −65.74; RPE (MS), −27.57; ANCCR, 18.60; Two-tailed p values = RPE (CSC), 1.7×10^{-83}, RPE (MS), 3.0×10^{-48}, ANCCR, 4.5×10^{-34}; n=100 simulations). **D.** Licking and dopamine response from two example mice (rewards with less than 3 s previous IRI were excluded to avoid confounding by ongoing licking responses). Though not our initial prediction, ANCCR can even account for the negative unpredicted sucrose response from Animal 2 (Fig S8). **E.** Quantification of correlation between dopamine response and number of rewards. Left panel shows the data from an example animal and the right panel shows the population summary across all animals (one sample t-test against a null of zero; t(7) = 4.40, two-tailed p = 0.0031; n=8 animals). Reward response was defined as the difference of area under curve (AUC) of fluorescence trace between reward and baseline period (Methods). **F.** Simulations confirming the intuitive reasoning from **B** for Test 2 (one sample t test against a null of zero; t(99) = RPE (CSC), -1.7×10^{3}, RPE (MS), −151.28, ANCCR, 335.03; Two-tailed p values = RPE (CSC), 5.0×10^{-223}, RPE (MS), 6.3×10^{-119}, ANCCR, 4.8×10^{-153}, n=100 iterations). **G.** Quantification of correlation between dopamine response and the previous IRI for an example session (left) and the population of all animals (one sample t-test against a null of zero; t(7) = 5.95, two-tailed p = 5.7×10^{-4}, n=8 animals). The average correlation across all sessions for each animal is plotted in the bar graph.

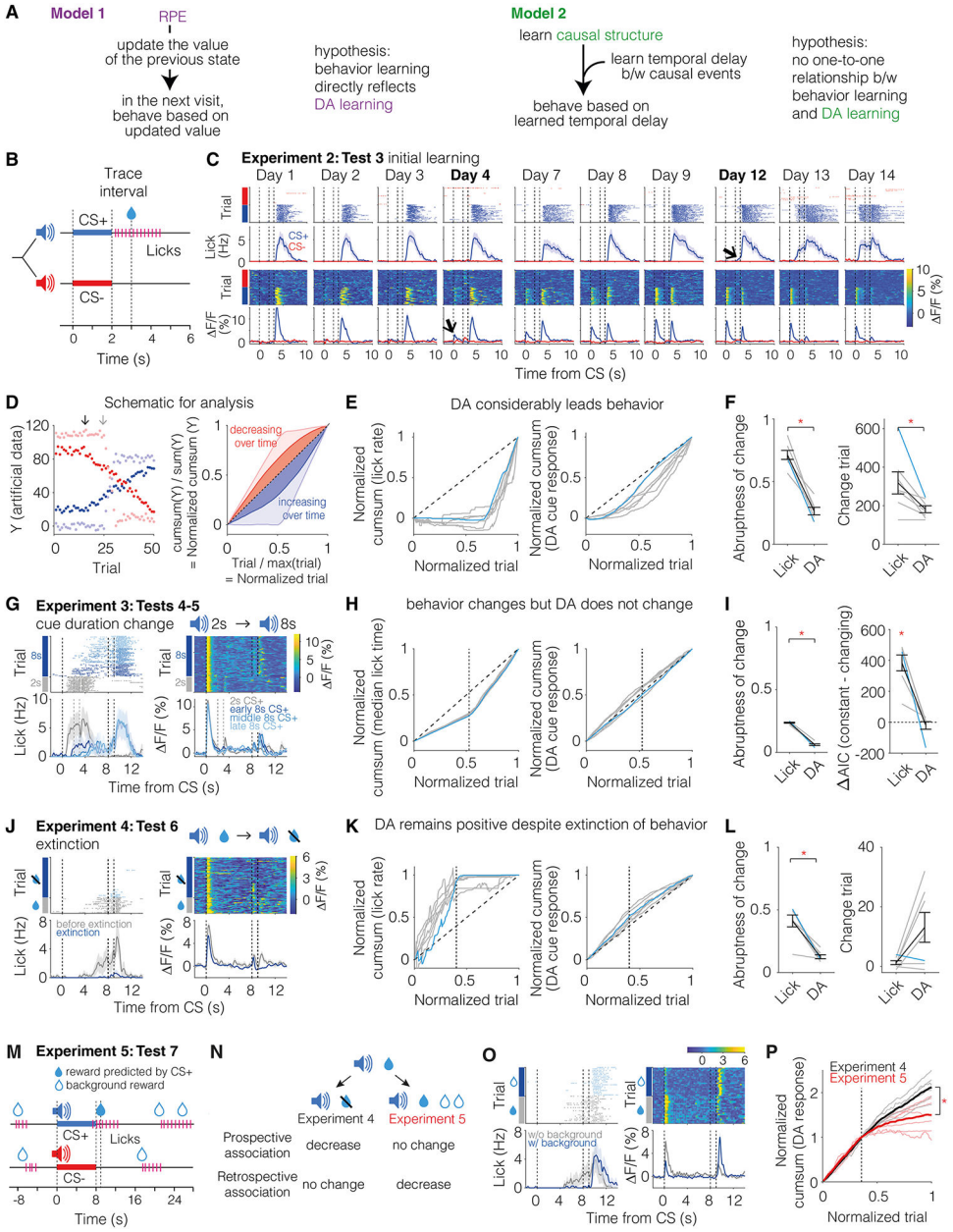

Fig. 4. The dynamics of dopamine responses during cue-reward learning are consistent with ANCCR, but not TDRL RPE.

A. TDRL predicts that dopaminergic and behavioral learning will be tightly linked during learning. However, the causal learning model proposes that there is no one-to-one relationship between behavioral and dopaminergic learning. **B.** Schematic of a cue-reward learning task in which one auditory tone predicted reward (labeled CS+) and another had no predicted outcome (labeled CS−). **C.** Licking and dopamine measurements from an example animal showing that the dopamine response to CS+ significantly precedes the emergence of anticipatory licking (Days 4 vs 12 respectively, shown by the arrows). **D.** Schematic to show a cumulative sum (cumsum) plot of artificial time-series data. A time-series that increases over trials appears below the diagonal in the cumsum plot with an increasing

slope over trials, and one that decreases over trials appears above the diagonal. Further, a sudden change in timeseries appears as a sudden change in slope in the cumsum plot. **E, F.** Dopamine cue response considerably leads behavior across animals. Each line is one animal, with the blue line corresponding to the example from **C.** Behavioral learning is much more abrupt than dopaminergic learning (paired t test for abruptness of change; $t(6) = 9.06$; two-tailed $p = 1.0\times10^{-4}$; paired t test for change trial; $t(6) = -2.93$; two-tailed $p = 0.0263$; n=7 animals). **G.** Anticipatory licking and dopamine release in an example animal after increasing the cue duration from 2 s to 8 s while maintaining a 1 s trace interval and a long ITI (~33 s). Trials are shown in chronological order from bottom to top. The three vertical dashed lines indicate cue onset, cue offset, and reward delivery (also in **J** and **O**). **H-I.** Behavior is learned abruptly by all animals, but the dopaminergic cue response shows little to no change. The dashed vertical line is the trial at which the experimental condition transitions (in **H, K**, and **P**). We tested for the lack of change by showing that the Akaike Information Criterion (AIC) is similar between a model assuming change and a model assuming no change. Paired t test for abruptness of change; $t(6) = 22.92$; two-tailed $p = 4.52\times10^{-7}$; one-sample t test for ΔAIC against a null of zero; $t(6) = 7.49$ for lick, -0.86 for dopamine; two-tailed $p = 2.9\times10^{-4}$ for lick, 0.4244 for dopamine (n=7 animals). **J.** The dopaminergic cue response of an example animal remains positive well after it learns extinction of the cue-reward association. **K-L.** Across all animals, the dopaminergic cue response remains significantly positive despite abrupt behavioral learning of extinction (paired t test for abruptness of change; $t(6) = 5.67$; two-tailed $p = 0.0013$; paired t test for change trial; $t(6) = -2.40$; two-tailed $p = 0.0531$; n=7 animals). **M.** Experiment to reduce retrospective association while maintaining prospective association. **N.** Two experiments that show specific reduction in either prospective or retrospective association. **O.** Licking and dopamine release from an example animal. **P.** Dopamine cue response reduces more rapidly during the background reward experiment in which the cue is followed consistently by a reward than during extinction in which there is no reward (paired t test; $t(6) = -3.51$; two-tailed $p = 0.0126$; n=7 animals).

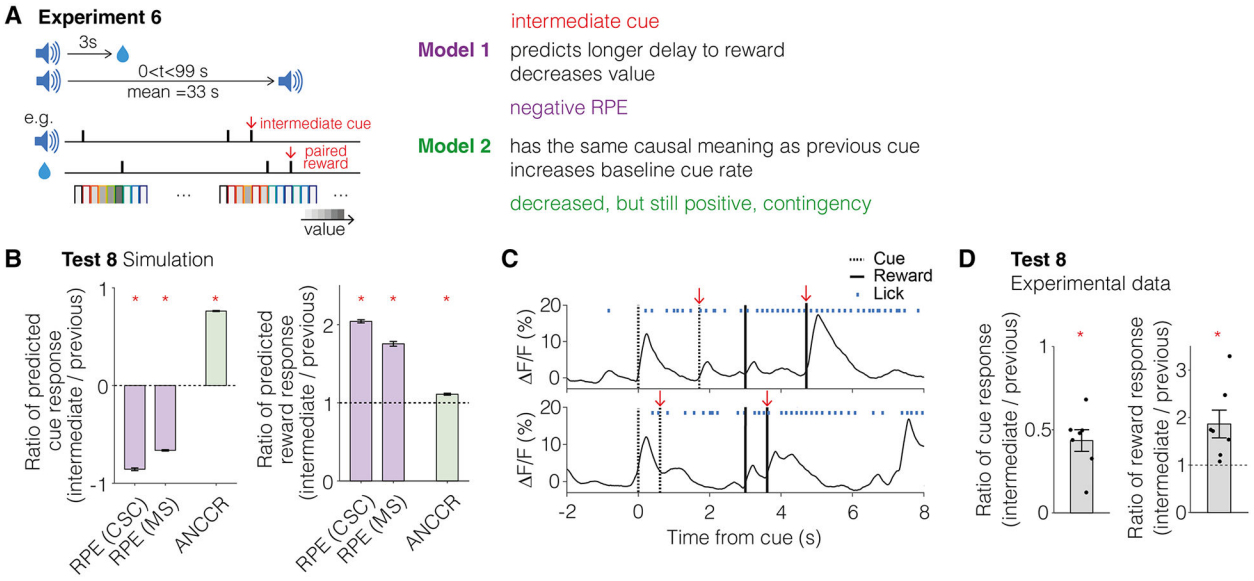

Fig. 5. Dopamine responses in a "trial-less" cue-reward task reflect causal structure like ANCCR, but unlike TDRL RPE.

A. A "trial-less" cue-reward learning task. Here, a cue (250 ms duration) is consistently followed by a reward at a fixed delay (3 s trace interval). However, the cues themselves occur with an exponential inter-cue interval with a 33 s mean. **B.** Confirmation of these intuitions based on simulations (Methods) (One sample t test against a null of zero; t(99) = RPE (CSC), −114.74; RPE (MS), −181.32; ANCCR, 322.53; Two-tailed p values = RPE (CSC), 4.1×10^{-107}; RPE (MS), 1.1×10^{-126}; ANCCR, 2.1×10^{-151}; n=100 iterations). Reward responses are predicted to be positive by both models (One sample t test against a null of one; t(99) = RPE (CSC), 87.67; RPE (MS), 62.86; ANCCR, 16.78; Two-tailed p values = RPE (CSC), 1.2×10^{-95}; RPE (MS), 1.3×10^{-81}; ANCCR, 1.1×10^{-30}; n=100 iterations). **C.** Example traces from one animal showing that the dopamine response to the intermediate cue is positive. **D.** Quantification of the experimentally observed ratio between the intermediate cue response and the previous cue response (One sample t test against a null of zero; t(6) = 6.64, two-tailed p value = 5.6×10^{-4}; n=7 animals), and reward response (One sample t test against a null of one; t(6) = 2.95; two-tailed p value = 0.0256; n=7 animals).

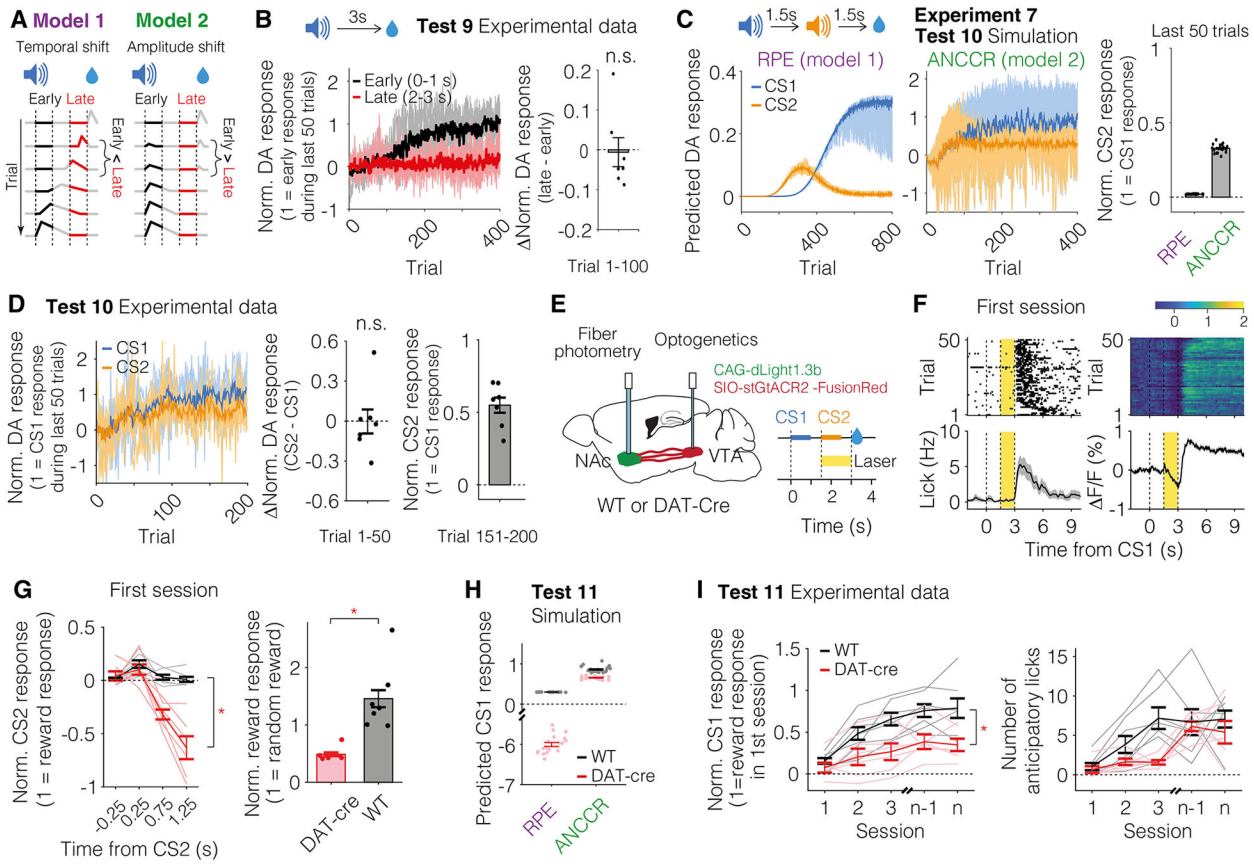

Fig. 6. No backpropagation of dopamine signals during learning.

A. Schematic of learning dynamics for pre-reward dopamine dynamics based on RPE or ANCCR signaling. Schematic was inspired from (50). If there is a temporal shift, the difference in dopamine response between early and late phases of a trial will be negative in the initial trials. **B.** Dynamics of dopamine response during early and late periods within a trial over training (left), and their difference during first 100 trials. **C.** Simulated dynamics for dopamine responses to cues (CS1 and CS2) during sequential conditioning (left), and averaged CS2 response during last 50 trials (right). **D.** Experimental data showing dynamics of dopamine responses to cues (left). Response difference between two cues during early phase of learning (middle; similar to Fig6B right) and CS2 response during late phase of learning (right, similar to Fig6C right). **E.** Schematic of optogenetic inhibition experiment during sequential conditioning for both experimental DAT-Cre animals receiving inhibition and control wild type animals receiving light but no inhibition. Animals received laser from CS2 until reward throughout conditioning. **F.** Measured licking and dopamine responses on the first session of conditioning from an example experimental animal, showing robust inhibition. **G.** Quantification of magnitude of inhibition during CS2 presentation prior to reward, and reward response. Both responses are measured relative to pre-CS1 baseline. **H.** Predicted dopamine responses using simulations of RPE or ANCCR. **I.** Experimental data showing CS1 response (left) and anticipatory licking (right) across sessions. Here, n represents the last session.